The Use of Latent Semantic Analysis in Operations Management Research

Kulkarni, Shailesh S.; Apte, Uday M.; Evangelopoulos, Nicholas E.

Kulkarni, Shailesh S., Uday M. Apte, and Nicholas E. Evangelopoulos. "The use of latent semantic analysis in operations management research." Decision Sciences 45.5 (2014): 971-994.

http://hdl.handle.net/10945/68511

DECISION SCIENCES
A JOURNAL OF THE DECISION SCIENCES INSTITUTE

Decision Sciences
Volume 45 Number 5
October 2014

The Use of Latent Semantic Analysis in Operations Management Research*

Shailesh S. Kulkarni†
Department of Information Technology and Decision Sciences, University of North Texas, 1307 W Highland, Denton, TX, 76201, e-mail: Shailesh.Kulkarni@unt.edu

Uday M. Apte
Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, CA 93943, e-mail: umapte@nps.edu

Nicholas E. Evangelopoulos
Department of Information Technology and Decision Sciences, University of North Texas, 1307 W Highland, Denton, TX, 76201, e-mail: Nick.Evangelopoulos@unt.edu

ABSTRACT

In this article, we introduce the use of Latent Semantic Analysis (LSA) as a technique for uncovering the intellectual structure of a discipline. LSA is an emerging quantitative method for content analysis that combines rigorous statistical techniques and scholarly judgment as it proceeds to extract and decipher key latent factors. We provide a stepwise explanation and illustration for implementing LSA. To demonstrate LSA's ability to uncover the intellectual structure of a discipline, we present a study of the field of Operations Management. We also discuss a number of potential applications of LSA to show how it can be used in empirical Operations Management research, specifically in areas that can benefit from analyzing large volumes of unstructured textual data.
[Submitted: August 16, 2012. Revised: January 17, 2014. Accepted: January 23, 2014.]

Subject Areas: Big Data Analytics, Latent Semantic Analysis, Operations Management Research, and and Unstructured Text.

INTRODUCTION

A characteristic development in today's economy is the volume and ubiquitous availability of structured as well as unstructured (textual) data—and its consideration as a resource for improved decision making. Unstructured data are available in many settings. Corporations regularly publish reports that contain information on the strategic vision/outlook of the company, financial outcomes, operations

*The authors would like to thank the Editor-in-Chief and the Associate Editor for their constructive comments which have improved the exposition of this article.

†Corresponding author.

strategy, and sustainability practices. Professionals and consumers take notes, produce reports, and offer testimonies containing critical observations that relate to their specific business. Customers submit online comments to company Web sites and social media that contain information related to the quality of products and services. *Latent Semantic Analysis* (LSA) is a powerful text analytics method that is capable of uncovering the conceptual content in unstructured data available in such business settings, and can be of significant benefit to both researchers and practitioners in the broad discipline of Decision Sciences.

Academics publish copious quantities of research articles that reflect the state of the art in their respective disciplines. On a broader scale, the combined content of such research articles reflects an underlying intellectual structure of the discipline, the understanding of which can help answer such questions as where the field has been, where its key publication outlets have shortcomings, what the contemporary trends in topics and research methods are, and what opportunities lie ahead. But how does one evaluate and assess the intellectual structure of a discipline? In the past, scholars have used thematic reviews, manual content analysis, and Citation–Cocitation Analysis (CCA). The goal of this article is to introduce LSA as a viable alternative that can be used for uncovering the intellectual structure of a discipline. For the benefit of the reader who is new to LSA, we also provide a small example that illustrates its underlying computational steps.

In this article, we propose LSA as a method that can match the insights provided by previously used methods and also provide interesting additional insights that are a direct outcome of LSA's treatment of comprehensive collections of unstructured text. LSA also addresses some shortcomings of alternative techniques by not relying on preconceived notions regarding the emerging themes and thereby limiting any subjective bias in the analysis. To demonstrate LSA's ability to uncover the intellectual structure of a discipline, we present a study of the field of Operations Management (OM). A similar study could, of course, be conducted to understand the intellectual structure of a functional area that comes under the purview of the Decision Sciences umbrella.

There have been several studies aimed at understanding the intellectual structure of various business disciplines. Marketing produced conceptual reviews as early as the 1940s (Applebaum, 1947). It later adopted manual content analysis starting with Kassarjian's (1977)) definitive work on consumer research and CCA (e.g., Hoffman & Holbrook, 1993). Management has also developed a tradition of taking stock of its intellectual structure with conceptual reviews (e.g., Boulding, 1958), manual content analysis (e.g., Brutus, Aguinis, & Wassmer, 2013), and CCA (Nerur, Rasheed, & Natarajan, 2008; Tsai & Wu, 2010). While still a very young discipline, Information Systems studied its intellectual structure using CCA (Culnan, 1986), following up with a number of conceptual reviews (e.g., Banker & Kauffman, 2004) and manual content analysis studies (e.g., Vessey, Ramesh, & Glass, 2002) and, eventually, LSA (Sidorova, Evangelopoulos, Valacich, & Ramakrishnan, 2008).

Past studies on the intellectual structure of OM research have primarily relied on conceptual reviews (e.g., Chopra, Lovejoy, & Yano, 2004), manual content analysis and categorization (e.g., Amoako-Gyampah & Meredith, 1989; Meredith, Raturi, Amoako-Gyanpah, & Kaplan, 1989; Filippini, 1997; Scudder & Hill, 1998;

Pannirselvam, Ferguson, Ash, & Siferd, 1999; Rungtusanthanam, Choi, Hollingworth, Wu, & Forza, 2003; Craighead & Meredith, 2008), or some variant of CCA (e.g., Pilkington & Liston-Heyes, 1999; Pilkington & Fitzgerald, 2006; Pilkington & Meredith, 2009).

Of the various methodological approaches available for studying a discipline's intellectual structure, the most quantitative and, at the same time, least subjective are CCA and LSA. While the strength of CCA for understanding the intellectual structure of a field is justifiably well recognized, also well known is its characterization as a "rear-view mirror" (White & McCain, 1998) due to its tendency to reflect the intellectual structure of a scientific field with considerable time lag. Moreover, because CCA relies on the researcher's understanding of cited works and not on the content of the original research articles or that of the cited works, it is an indirect method for the delineation of the intellectual structure of a field. In contrast, as a text analysis method, LSA has the ability to directly study the content of research articles without causing any time delay. Furthermore, LSA can provide the same insights as qualitative techniques such as manual content analysis and conceptual reviews but, due to its reasonably sophisticated quantitative engine, it is capable of surfacing other insights that may be very difficult or impossible to glean from these other techniques. In summary, with LSA, one can expect to reach sound and credible conclusions regarding the intellectual structure of a discipline and estimate the timing of its trends with better accuracy.

The remainder of this article is organized as follows: The section "Latent Semantic Analysis" provides a self-contained primer on LSA. Section "The state of OM research: an LSA application" presents a detailed application of LSA to OM research and discusses our main findings related to the intellectual structure of OM. It is important, however, to note that LSA's domain of application extends beyond analyzing published research articles and includes other kinds of unstructured text data that may be encountered by researchers. To this end, the article concludes in the section "Directions for future applications of LSA" by discussing a number of potential applications of LSA to other research problems.

LATENT SEMANTIC ANALYSIS

LSA is a statistical method for estimating the meaning in words and passages as linear combinations of underlying concepts. These underlying concepts are extracted through matrix operations of observed patterns of word usage. The fundamental idea behind LSA is that the meaning of each passage of text (a "document") is modeled as the sum of meanings of the individual words in it, whereas a collection of documents (a "corpus") is modeled as a system of simultaneous equations that can determine the similarity in meaning of words and documents to each other.

LSA originated in the late 1980s/early 1990s (Deerwester Dumais, Furnas, Landauer, & Harshman, 1990) as an information retrieval technique designed to improve library indexing and search engine query performance (see, e.g., Dumais, 2004; Han & Kamber, 2006; Cios, Pedrycz, Swiniarsk, & Kurgan, 2007; Dumais, 2007; Manning, Raghavan, & Schütze, 2008).

The cognitive psychology theory behind LSA is intimately linked to the acquisition, induction, and representation of knowledge. A simple mechanism of induction, based on a high-dimensional linear associative model, provides an explanation of why people know more than they should, simply based on the individual pieces of information to which they have been exposed (Landauer & Dumais, 1997). LSA provides a mathematical model for this induction mechanism.

LSA has been applied by psychology researchers as a theory and method for extracting and representing the meaning of words by humans, including word sorting and category judgments (Landauer, 2007). In recent years, LSA has been applied to the fields of information retrieval, artificial intelligence, psychology, cognitive science, education, information systems, and many others.

Martin and Berry (2007) provide an introduction to the mathematics of LSA and a small numerical example that illustrates how the analysis works. Valle-Lisboa and Mizraji (2007) provide a rigorous discussion on how LSA detects the underlying topical structure of a document corpus and why LSA's capability for discovering hidden topics allows it to successfully model synonyms, multiple words with similar meaning, and human memory. The specific steps in using LSA are listed below.

Step 1: Compilation of Term Frequency Matrix (Vector Space Model [VSM])

LSA starts with a text quantification method known as the VSM (Salton, 1975), where a corpus of d documents using a vocabulary of t terms is used to compile a $t \times d$ matrix $\mathbf{X}$, containing the number of times each term appears in each document (term frequencies). Before the term cardinality is finalized, a number of term filtering or term consolidation (term selection) operations are performed. Some frequent, but trivial, terms such as "the," "of," (the *stoplist*) are excluded because they carry little information. Some infrequent terms are also excluded because they play too small a part in shaping the corpus, and terms sharing a common stem are consolidated (*term stemming*, Porter, 1980).

The frequency counts in $\mathbf{X}$ typically undergo some transformation (*term weighting*) that penalizes common terms and promotes less common ones. After weighting, the term frequencies are typically also normalized so that the sum of squared transformed frequencies of all term occurrences within each document is equal to one (Salton & Buckley, 1988).

Step 2: Singular Value Decomposition [SVD]

The reduced and transformed version of the term frequency matrix, $\mathbf{A}$, is subjected to SVD, $\mathbf{A} = \mathbf{U\Sigma V}^{\mathrm{T}}$, where $\mathbf{U}$ are the term eigenvectors, $\mathbf{V}$ are the document eigenvectors, $\mathbf{\Sigma}$ is a diagonal matrix of singular values (i.e., square roots of common eigenvalues between terms and documents in the least-squares sense), and the superscript $^{\mathrm{T}}$ denotes transposition.

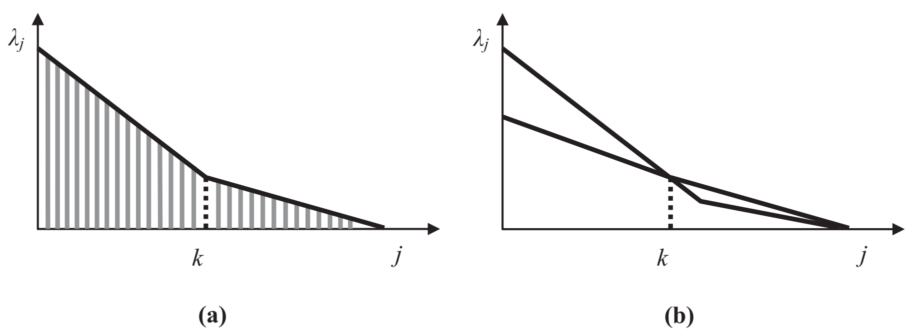
Step 3: Dimensionality Selection and Factor Rotations

The problem of selecting an optimal number of latent semantic dimensions is typically dealt with empirically (Bradford, 2008 and see Table 1 for a tabulation of 49 studies with optimal factor numbers ranging from 6 to over 1,000) and remains open. Certain quantitative approaches have also been proposed recently. In our

Table 1: Titles of five selected articles published in *POM*.

ID	Article title	*POM* Reference
D1	Dynamic Procurement, Quantity Discounts, and **Supply Chain** Efficiency	17(5), 543–550
D2	Coordinating a **Supply Chain System** with Retailers **Under** Both Price and **Inventory** Competition	17(5), 532–542
D3	Multiperiod Models with Capacities in Competitive **Supply Chain**	17(4), 439–454
D4	Strategic **Management** of Distressed **Inventory**	17(4), 402–415
D5	Contracting **Under** Vendor **Managed Inventory Systems** Using Holding Cost Subsidies	17(2), 200–210

Figure 1: Dimensionality estimation in LSA: (a) based on LLR, (b) based on parallel analysis.



study, we follow the log-likelihood ratio (LLR) test approach (Zhu & Ghodsi, 2006). Conceptually, the LLR approach seeks to quantitatively estimate an "elbow point" on the eigenvalues scree plot, k, as shown in Figure 1(a).

This elbow point is the point of diminishing returns for the eigenvalues' ability to explain variance in the term frequency matrix $\mathbf{A}$. Another approach that is not pursued here is amended parallel analysis (Efron, 2005), which is a bootstrap-based comparison of simulated sets of eigenvalues under the null hypothesis of term independence versus the observed set of eigenvalues on the scree plot, as shown in Figure 1(b). Keeping the first k dimensions produced by SVD, a reconstruction operation can now produce $\hat{\mathbf{A}}_k$, a least-squares approximation of $\mathbf{A}$, which encompasses information from the first k principal components of the corpus space:

$$\mathbf{A}_k = \hat{\mathbf{U}}_k \Sigma_k \mathbf{V}_k^{\mathrm{T}}. \tag{1}$$

In the factor analysis variant of LSA, the factor space may be transformed to a new base by rotating the term loadings, thus allowing the researcher to interpret the latent semantic factors (Sidorova et al., 2008; Evangelopoulos, Zhang, & Prybutok, 2012). The term variance–covariance matrix $\hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathrm{T}}$ can be reproduced using the unrotated term loadings $\mathbf{U}_k\mathbf{\Sigma}_k(\mathbf{U}_k\mathbf{\Sigma}_k)^{\mathrm{T}}$, or the rotated term loadings $\mathbf{U}_k\mathbf{\Sigma}_k\mathbf{M}_k(\mathbf{U}_k\mathbf{\Sigma}_k\mathbf{M}_k)^{\mathrm{T}}$, where $\mathbf{M}_k$ is any matrix that satisfies the orthonormality property $\mathbf{M}_k\mathbf{M}_k{}^{\mathrm{T}} = \mathbf{I}_k$, $\mathbf{I}_k$ being the identity matrix of rank k. $\mathbf{M}_k$ can be computed as the rotation matrix by any factor rotation procedure, such as varimax. Due to the varimax procedure's capability of simplifying the term loadings by making them either very large or very small, the rotated factor space can be easily interpreted by associating the factors with high-loading terms obtained from the rotated term loadings $\mathbf{U}_k\mathbf{\Sigma}_k\mathbf{M}$ and high-loading documents obtained from the rotated document loadings $\mathbf{V}_k\mathbf{\Sigma}_k\mathbf{M}$.

Step 4: Factor Labeling

To the extent that the extracted LSA factors represent a set of topics that characterize the domain of analysis, rotated factors can be interpreted and labeled. In this step, one first looks at the high-loading terms to surmise what concepts the factor in question is likely referring to, then takes a look at the high-loading documents to better conjecture what the factor is, and finally gives a suitable label to the factor using common terms in the field. It goes without saying that being well versed in the field is critically important for arriving at correct interpretation and meaningful labeling of factors.

Post-LSA Analysis

Once the latent semantic factors have been extracted, they can be used for further quantitative analysis in conjunction with existing document attributes such as document source (association analysis) or document time stamp (time trends). Furthermore, new documents, often referred to as *query* documents, can be represented in the factor space. Considering a collection of m query documents expressed in the original space of t terms as the $t \times m$ pseudo-documents matrix $\mathbf{Q}$, its representation in the factor space can then be obtained by the *query loadings* matrix $\mathbf{L}_{\mathrm{Q}} = \mathbf{Q}^T\mathbf{U}_k$, where $\mathbf{U}_k$ is obtained from (2). In the rotated factor space, after the rotation matrix $\mathbf{M}_k$ is obtained through a rotation procedure such as varimax, query loadings can be obtained as

$$\mathbf{L}_{\mathrm{Q}}\mathbf{M}_k = \mathbf{Q}^{\mathrm{T}}\mathrm{U}_{\mathrm{k}}\mathbf{M}_{\mathrm{k}}. \tag{2}$$

An association between a reference factor space and new documents can then be made based on these new document loadings. Representing new documents in this manner, without re-computing the factor space, is known as *fold-in*. Fold-in has the disadvantage of failing to detect new concepts that were not previously extracted in the old factor space, but has the advantage of maximum computational efficiency. Fold-in can be an excellent methodological choice in environments where the concept space changes very slowly, but incoming documents are added to the data set at high rates that overwhelm the capacity to process concepts using approaches that require either human interpretation, or continuous recomputation. For more discussion on the fold-in approach and its comparison to SVD re-computation see Berry Dumais, and O'Brien (1995).

LSA Implementation Software

Commercial software packages that implement LSA include the proprietary SAS Text Miner, which currently offers free access to academics through the SAS OnDemand for Academics program (http://www.sas.com/govedu/edu/programs/od_academics.html). Open-access software options include the R package LSA (http://cran.r-project.org/web/packages/lsa/index.html), which requires some customization in order to perform rotations. The computations presented here are based on the Java package JAMA, offered by the National Institute of Standards and Technology (http://math.nist.gov/javanumerics/), which requires a text mining front-end addition, such as JTMT (http://jtmt.sourceforge.net/) and some customization in order to perform rotations.

Brief Illustrative Example of LSA Computations

For the benefit of the reader who is unfamiliar with LSA, a small numerical example that illustrates associated computations, as well as underlying premises, follows. Our illustrative example is based on a very small set of textual data (shown in Table 1) and follows the numbered steps presented earlier. The documents are titles of five selected articles published in *Production and Operations Management* (*POM*) in 2008 (vol. 17).

Step 1: The initial vocabulary of 35 raw terms is reduced through the following term selection methods: (i) the application of a *stoplist* (i.e., the exclusion of trivial English words such as "and" or "of"), which reduced the terms to 28; (ii) the application of term stemming (i.e., the replacement of terms with a common stem by that stem—e.g., "managed" and "management" are consolidated as "manag-," and "system" and "systems" get consolidated as "system-"), reducing the vocabulary to 26 stemmed terms; and (iii) the exclusion of terms that appear only once in the document collection (e.g., "capacities" or "procurement"). These term selection methods produce the final vocabulary of six stemmed terms: {*chain*, *inventori-*, *manag-*, *suppli-*, *system-*, *under*}. Table 1 identifies the occurrence of original words related to these six terms by boldfacing.

Table 2 shows the raw term frequencies for each of the five documents, organized in a 6×5 term-by-document matrix. Table 3 shows the term frequency matrix after a transformation based on Inverse Document Frequencies (TF-IDFs transformation), which penalizes frequent terms and promotes rare terms (Salton, 1975; Robertson, 2004; Sidorova et al., 2008; Wei, Hu, Tai, Huang, & Yang 2008) and normalization, so that the sum of squared frequencies in each column is equal to 1.

TF-IDF multiplies local (i.e., pertinent to the particular term in the particular document) term frequency (TF) by global (i.e., pertinent to the entire collection of documents) IDF. More specifically, the TF-IDF transformation as used in our analysis replaces the raw term frequency TF_{ij} of term i in document j by

$$w_{ij} = TF_{ij}IDF_i = TF_{ij}\log_2(N/n_i), \tag{3}$$

Table 2: Raw term frequencies for the titles in Table 1, organized as a 6 × 5 matrix.

	Document				
Term	*D1*	*D2*	*D3*	*D4*	*D5*
chain	1	1	1	0	0
inventori-	0	1	0	1	1
manag-	0	0	0	1	1
suppli-	1	1	1	0	0
system-	0	1	0	0	1
under	0	1	0	0	1

Table 3: Transformed term frequencies after term frequency inverse document frequency (TF-IDF) weighting and normalization.

	Document				
Term	*D1*	*D2*	*D3*	*D4*	*D5*
chain	0.707	0.326	0.707	0	0
inventori-	0	0.326	0	0.487	0.306
manag-	0	0	0	0.873	0.550
suppli-	0.707	0.326	0.707	0	0
system-	0	0.584	0	0	0.550
under	0	0.584	0	0	0.550

where N is the total number of documents in the collection and n_i the frequency of term i in the entire collection of documents.

Step 2: The matrix shown in Table 3 is subjected to SVD.

Step 3: Figure 2 shows a scree plot of the five eigenvalues (2.47, 1.78, 0.72, 0.03, and 0) produced by this analysis. Note the rank deficiency of our matrix resulting in the last eigenvalue being zero, due to the identical representation of documents *D1* and *D3* in the term frequency matrix in Table 3. Based on this plot, keeping the first two principal components seems appropriate, because they are both larger than the mean eigenvalue, which is equal to 1. Note that, while the application of this criterion (known in factor analysis as the Kaiser–Guttman rule) for the selection of an appropriate number of factors in this minimalistic example produces factors that describe well the content of the five titles, the presence of a much larger vocabulary makes similar rules less applicable.

Step 4: Interpretation and labeling of the two extracted factors is the next step in our analysis. Table 4 shows the term loadings before and after a varimax rotation. Factor F1 appears to be mostly related to terms {*chain*, *suppli-*},

Figure 2: Scree plot.

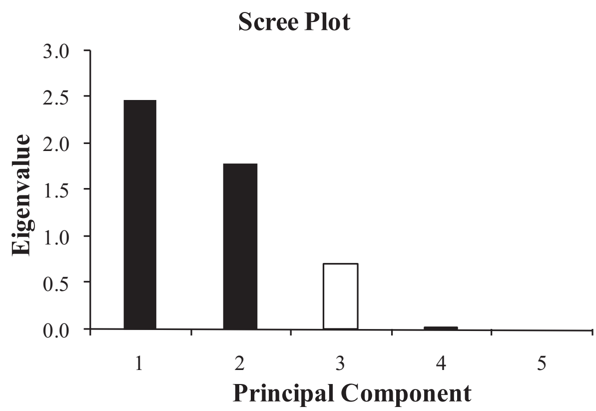


Table 4: Term loadings before and after varimax rotation.

	Unrotated		Rotated	
Term	F1	F2	F1	F2
chain	−0.911	0.515	**−1.043**	−0.092
inventori-	−0.386	−0.499	−0.034	**−0.630**
manag-	−0.381	−0.783	0.130	**−0.862**
suppli-	−0.911	0.515	**−1.043**	−0.092
system-	−0.507	−0.441	−0.168	**−0.651**
under	−0.507	−0.441	−0.168	**−0.651**

whereas factor F2 appears to be primarily related to terms {*inventori-*, *manag-*, *system-*, *under*}. Table 5 shows the document loadings before and after the same varimax rotation that was applied to the term loadings (i.e., using the same rotation matrix). Factor F1 loads high on documents *D1*, *D2*, and *D3*. Factor F2 loads high on documents *D2*, *D4*, and *D5* (note the cross-loading of document *D2*). Reading again the corresponding titles from Table 1, it is plausible to infer that factor F1 is about Supply Chains and factor F2 is about Inventory Management Systems.

In order to better understand how terms and documents are represented in the latent semantic factor space, let us now examine how term frequencies are approximated by reconstructing the term frequency matrix after retaining the first two principal components (see Table 6). Using this two-factor space, the term frequencies appear modified from their original values in Table 3. For example, even though "*system-*" did not appear at all in document *D4* (see Table 3, column 4), it now does, and its frequency is quite high (see Table 6, column 4, highlighted cell). After examining this term-document structure and considering the statistical

Table 5: Document loadings before and after varimax rotation.

	Unrotated		Rotated	
Document	F1	F2	F1	F2
D1	−0.820	0.546	**−0.985**	−0.016
D2	−0.835	−0.256	**−0.542**	**−0.684**
D3	−0.820	0.546	**−0.985**	−0.016
D4	−0.332	−0.694	0.121	**−0.760**
D5	−0.564	−0.800	−0.010	**−0.979**

Table 6: Fitted term frequencies, produced using the first two principal components.

	Document				
Term	*D1*	*D2*	*D3*	*D4*	*D5*
chain	0.686	0.385	0.686	−0.076	0.018
inventori-	−0.003	0.301	−0.003	0.341	0.437
manag-	−0.121	0.355	−0.121	0.488	0.606
suppli-	0.686	0.385	0.686	−0.076	0.018
system-	0.084	0.354	0.084	0.337	0.446
under	0.084	0.354	0.084	0.337	0.446

patterns that are represented by the first two latent semantic factors, our LSA model suggests that when document *D4* mentions "inventory" and "management," it must be also talking about "systems." The cognitive science and psychology literature have proposed that this approximation imitates the way our human brain learns and draws conclusions.

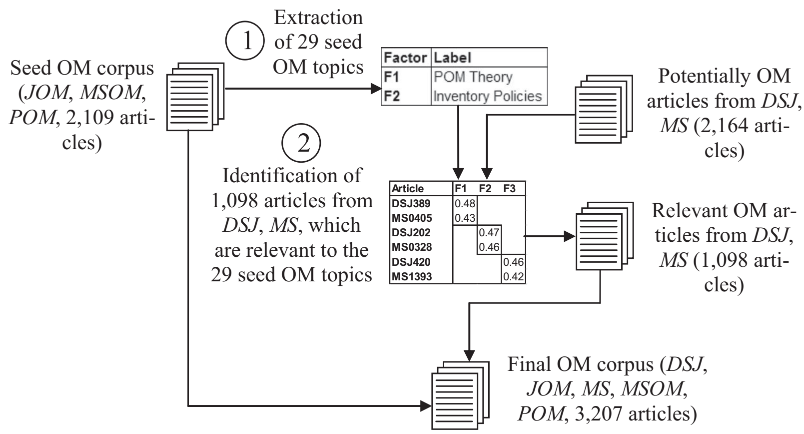
THE STATE OF OM RESEARCH: AN LSA APPLICATION

In this section, we use LSA to discern the intellectual structure of OM research. The details of our application appear below.

Data Collection

As the primary data in our study, we use abstracts of OM research articles published in five research journals, *Decision Sciences Journal* (*DSJ*), *Journal of Operations Management* (*JOM*), *Management Science* (*MS*), *Manufacturing and Service Operations Management* (*MSOM*), and *POM*. These journals are known for publishing high-quality OM research and combine long history (*DSJ*, *MS*, and *JOM* were already in print in 1980), affiliation with professional societies (Decision Science Institute for *DSJ*, American Production and Inventory Control Society for *JOM*, Institute for Operations Research and Management Sciences for MS

Figure 3: Steps in identification of our final OM corpus.



and MSOM, Production and Operations Management Society for *POM*), and consistent appearances in journal ranking lists. Given the interdisciplinary nature of *DSJ* and *MS*, we follow a very specific approach for identifying the subset of articles that are OM related from these two journals. These pre-LSA data preparation steps are shown in Figure 3. Abstracts of all research papers published in *JOM*, *MSOM*, and *POM* in the period 1980–2012 were obtained from the electronic library EBSCO to provide a starting OM corpus. Excluding editorial notes, book reviews, and commentaries, this seed corpus included 2,109 abstracts. Initial LSA application following the steps outlined in the previous section extracted 29 seed OM topics (see Figure 3, Step 1). Using the fold-in approach described in the previous section, the abstracts of 2,164 potentially OM articles (based on accepting department for MS and including all articles in *DSJ*) from *DSJ* and *MS* were related to the 29 seed topics. This step resulted in the identification of 1,098 OM-related articles published in DSJ and MS (Figure 3, Step 2), which were then added to the seed corpus to generate the final OM corpus of 3,207 articles published in *DSJ*, *JOM*, *MS*, *MSOM*, and *POM*. The loading threshold was identified based on the heuristic that all abstracts should, on average, load on one topic. The same heuristic was used in Sidorova et al. (2008) and Evangelopoulos et al. (2012). A breakdown of the final set of 3,207 articles is presented in Table 7.

Extraction of OM Research Topics

As mentioned earlier, this application study seeks to understand and analyze the major trends in OM research published in mainstream, well-respected OM journals during the past three decades (1980–2012). Thus, the primary goal is to identify the core research topics that OM scholars have focused on in the past three decades

Table 7: Research journals and OM articles included in our study.

Journal	1980–1989	1990–1999	2000–2009	2010–2012	Total
DSJ[a]	162	206	151	50	569[a]
JOM	207	232	409	127	975
MS[a]	135	156	180	58	529[a]
MSOM	n/a[b]	10[b]	263	118	391
POM	n/a[b]	193[b]	370	180	743
Total	504	797	1339	533	3,207

[a]Only OM-related articles from *DSJ* and *MS* were included in this step of the analysis.
[b]*MSOM* and *POM* were launched in 1998 and 1992, respectively.

and, thereby, understand the intellectual structure of OM research. Specifically, we identify the research topics studied during (i) the entire 1980–2012 period; (ii) the three individual decades of the 1980s, 1990s, and 2000s; and (iii) during the last 3 years in the current decade (2010–2012).

Analyzing both shorter and longer time frames illustrates the value of LSA in being capable of providing rich insights. From a technical standpoint, performing the analysis over a longer period of time offers the opportunity to extract factors that are more stable, and that have survived the test of time. On the other hand, extending the time frame too much may result in a nonuniform sample, where the way researchers are using language has undergone paradigmatic changes. From this point of view, shorter time frames (but not too short) within which article authors have functioned as a community, are also valuable.

OM Research Topics during 1980–2012

The steps followed in conducting LSA of the final OM corpus are shown in Figure 4. To identify research topics from OM abstracts, we first developed a vocabulary of 1,078 stemmed terms using the method described in the earlier example. This vocabulary allowed us to prepare a 1,078-term by 3,207-document matrix representing our OM corpus (Figure 4, Step 1). In order to compress the range of term frequencies, we applied the TF-IDF transformation as shown in Equation (3). The second step was to perform SVD as shown in Equation (1) (Figure 4, Step 2). As a third step, we selected a dimensionality and performed factor rotations (Figure 4, Step 3). The end products of the third step were sets of high-loading terms and articles for each of the extracted topic factors. These were used in labeling the topics (Figure 4, Step 4).

In studying the OM research published in the entire 1980–2012 period, the first question that we needed to resolve was the level of granularity to be used in the analysis, that is, deciding if the field is to be studied at a high level leading to a handful of broad research topics, or if it is to be studied at a very detailed level resulting in a large number of narrowly defined topics. To the extent that we want to understand the major trends in OM research, we chose to follow the former approach. As mentioned earlier, the problem of determining the optimal number of factors in LSA remains open and is typically addressed empirically through trial

Figure 4: Steps in main Latent Semantic Analysis of our final OM corpus.

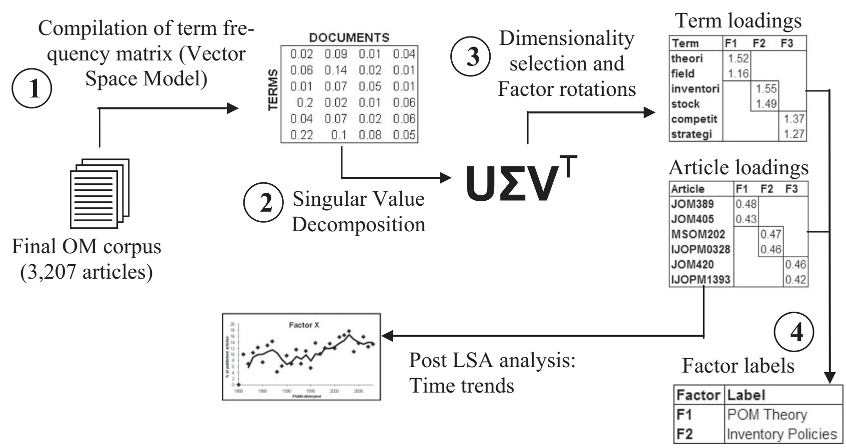


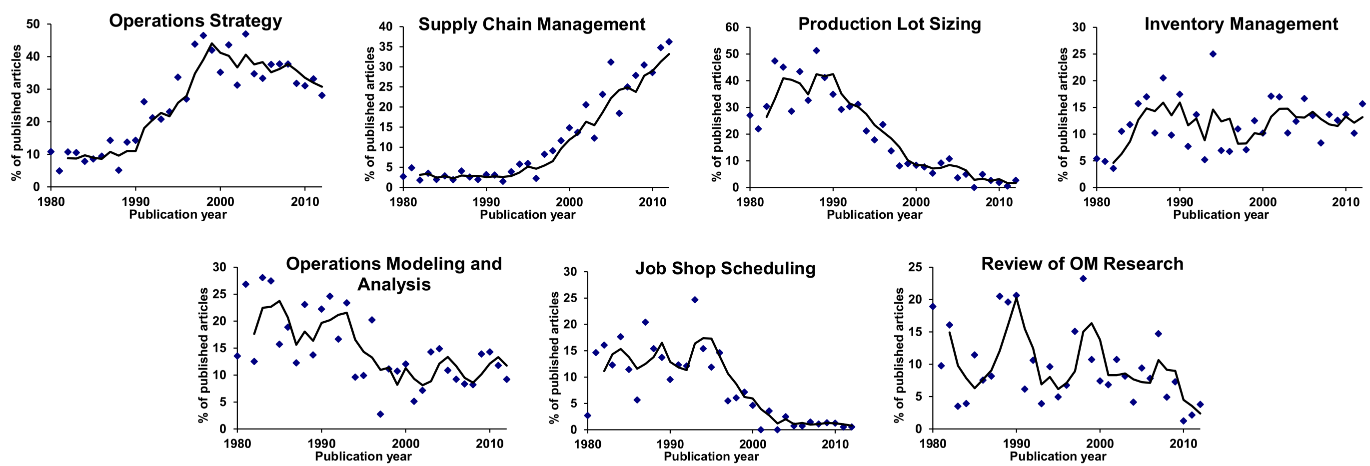
Table 8: High-level view of OM research topics during 1980–2012.

Factor No.	Factor Label	# Articles	% Var. Explained
1	Operations strategy	969	24.11
2	Supply chain management	551	16.30
3	Production lot sizing	429	15.34
4	Inventory management	388	11.95
5	Operations modeling and analysis	415	11.45
6	Job shop scheduling	183	11.00
7	Review of OM research	275	9.85

and error. Certain quantitative approaches such as the LLR test approach (Zhu & Ghodsi, 2006) are also available. For the high-level topics from the entire 1980–2012 period, however, because the LLR test cannot operate on samples of fewer than 10 dimensions, we used an empirical approach involving multiple trials of LSA with the number of factors in individual trials ranging from 2 to 10. After reviewing titles and abstracts of high-loading articles for each factor within each trial, we settled on seven factors as representative of core, high-level research areas within OM. See Table 8 for a listing of seven latent factors labeled independently by two coauthors. Table 8 also shows, for each topic, the corresponding article count, as well as percentage of variance explained. Finally, annual counts of high-loading articles were used to generate time trend plots for each factor. A complete listing of the seven trend plots over the past three decades is presented in Figure 5.

The seven-factor solution to the analysis of OM topics illustrates the power of LSA to extract latent concepts. Some obvious and expected observations from the trend plots in Figure 5 are the decline of research activity in areas in which a considerable amount of intellectual capital was invested in the early decades (e.g., areas such as Job Shop Scheduling and Production Lot Sizing). LSA also

Figure 5: Trends in high-level OM research topics during 1980–2012.



correctly indicates that the study of the topic of Supply Chain Management (SCM) has experienced a dramatic increase and sustained interest during the past two decades. Finally, an examination of the seven-factor solution indicates that Operations Strategy remains a dominant topic but academic interest in it seems to be on the decline in recent years.

Decade-Wise Analysis of OM Research Topics

Our LSA model mathematically represents discourse on OM research at any given point in time using contexts and word associations from the applicable observation period. To examine the dynamic behavior of the intellectual structure in OM, we also examine each decade separately. To this end, we used the LLR test approach to perform independent rounds of LSA of the 504 abstracts published in 1980–1989, the 797 abstracts published in 1990–1999, and the 1,339 abstracts published in 2000–2009. Topics extracted from these three decades are presented in Table 9.

In Table 9, a 14-factor analysis for the 1980s reveals, among others, the presence of Assembly Line Balancing, Job Shop Scheduling, Dynamic Lot Sizing, Facility Location, and MRP. A 17-factor analysis of the 1990s reveals the advent of JIT, AMT, Cellular Manufacturing, and TQM, and more importantly, the first emergence of factors closely related to SCM such as Supplier Management and Pricing besides of course the prevalent Multi-Echelon Inventory management. An interesting factor in the 1990s is OM Theory, which we believe captures the paradigmatic debate in the 1990s about what constituted OM research.

A 24-factor analysis for the most recent decade reveals some significant consolidation of topics. In fact, several of these topics, for example, Supply Chain Coordination, Auctions and Procurement, Buyer–Supplier Relationships, and even Remanufacturing relate to specialized pockets of SCM, thereby establishing its dominance as the most researched OM topic of the late 1990s and 2000s. The 2000s also saw the advent of E-Business and RFID, and Call Center Operations, thereby signaling specific business trends that OM scholars were quick to investigate; this is captured efficiently by our LSA.

OM Research Topics: A 2010–2012 Update

As a follow-up on the decade-wise analysis, we analyzed the 533 abstracts published in 2010–2012 to extract 14 topics of current interest to OM scholars. A list of these 14 topics is presented in Table 10 (right panel). This analysis, unavoidably, suffers the limitation of a relatively small sample size. Nevertheless, it presents useful information by identifying emerging research topics enjoying significant current interest. SCM is clearly still the dominant research area.

Interest in topics such as Global Operations and Healthcare Operations appears to be strengthening while Flexibility appears to be the new topic of interest during the past 3 years. We do not, however, see the erstwhile topic of Environmentally Responsible Operations from the previous decade being currently represented. Perhaps this is due to the small sample size in the analysis. The results of the new analysis of the 533 abstracts were compared to a fold-in of these abstracts on the previously extracted 24 factors from the 2000–2009 period, using Equation (2). See Table 10, left panel. Interestingly, the fold-in approach was able to identify

Table 9: Decade-wise analysis of OM research topics: 1980s, 1990s, and 2000s.

1980–1989 (14 topics)	1990–1999 (17 topics)	2000–2009 (24 topics)
Assembly Line Balancing	Advanced Manufacturing Technology (AMT)	Assembly Systems
Dynamic Lot Sizing	Cellular Manufacturing	Auctions and Procurement
Facility Location	Forecasting	Buyer–Supplier Relationships
Forecasting	JIT/Setup Reduction	Call Center Operations
Goal Programming	Job Shop Scheduling	Capacity Invest. & Risk Pooling
Job Shop Scheduling	Lot Sizing and Scheduling	E-Business and RFID
Learning Curve Analysis	Manufacturing Flexibility	Environmentally Responsible Operations
Maintenance Management	Multi-Echelon Inventory	Forecasting
Manufacturing Strategy	Operations Management Pedagogy	Knowledge Management
Material Requirements Planning	Operations Management Theory	Lot-Sizing
Multi-Echelon Inventory	Pricing in Operations	Manufacturing Flexibility
Project Scheduling	Production Planning and Control	Multi-Echelon Inventory
Quantity Discounts and Pricing	Project Scheduling	Operations Integration
Safety Stocks in Inventory	Service Operations	Outsourcing and Offshoring
	Supplier Management	Pricing in Supply Chains
	Total Quality Management	Project Management
	Workforce/Labor Flexibility	Remanufacturing
		Resource Network Design
		Review of OM Research
		Service Operations
		Service System Design
		Supply Chain Coordination
		Total Quality Management
		Workforce Management

10 articles published in 2010–2012 that relate to the previously extracted Environmentally Responsible Operations topic. This illustrates the trade-off between the identification of previously unknown topics in a new analysis, and the efficient, fully automated identification of new documents that relate to previously known topics in fold-in analysis.

Table 10: OM topics in 2010–2012: Fold-in versus new LSA extraction.

Fold-In Using the 24 Topics in 2000–2009		Extraction of New Topics in 2010–2012	
Topic from 2000–2009	No. of 2010–2012 articles	Topic in 2010–2012	No. of 2010–2012 articles
Pricing in Supply Chains	80	Buyer–Supplier Relationships	52
Operations Integration	56	Inventory and Replenishment	50
Multi-Echelon Inventory	30	Pricing & Selling in Supply Chains	50
Buyer–Supplier Relationships	28	Supply Chain Coordination	50
Supply Chain Coordination	25	Supply Chain Contracting	48
Project Management	20	Revenue Management	38
Call Center Operations	16	E-Business & Retail Supply Chains	37
Lot-Sizing	15	Healthcare Operations	35
Capacity Investment and Risk Pooling	13	Operations Integration	32
Knowledge Management	13	Project Management	30
Resource Network Design	13	Global Operations	25
Forecasting	12	Flexibility	23
Manufacturing Flexibility	12	Call Center Operations	22
Remanufacturing	12	Remanufacturing	14
Service System Design	12		
Outsourcing and Offshoring	11		
Environmentally Responsible Operations	10		
Service Operations	10		
Auctions and Procurement	8		
Workforce Management	8		
Assembly Systems	7		
Review of OM Research	7		
E-Business and RFID	6		
Total Quality Management	5		

Limitations and Potential Extensions of the Study

There are limitations to the LSA study of OM research presented in this section. First, our analysis, although grounded in sound quantitative techniques, is still somewhat subjective. Deciding on the optimal number of factors, cut-offs for factor loadings and, more important, the naming of the latent factors themselves requires trial-and-error and some human judgment. Second, when the time period covered by the data analysis is somewhat short, such as in the 2010–2012 analysis, there is no guarantee that the LSA will always uncover all important and emerging topics in the field, simply because the number of articles published on some of those topics will perhaps be not large enough to warrant their inclusion. This is a broader limitation of any study such as ours due to the long gestation periods of academic research in terms of concept formulation to commissioning.

We see several avenues for extending the application study presented in this section. One obvious extension is to study OM research at a detailed level, resulting in identification of 30 or 40 narrowly defined research topics. Our study could also be replicated in the future using the latest developments in LSA or its successors. Such a study will benefit from the inclusion of relatively longer time periods for POM and MSOM, assuming that a 20- to 30-year time frame is considered. To some extent, such future work will ascertain the "moving-average" of the OM core. Another avenue for future work is the inclusion of several other OM-related journals in an LSA of topical emphases. One could also analyze the ancillary journals not considered in this analysis to find out what part of the OM core is covered by these journals. Finally, it would be interesting to find out how the journals position themselves when represented in the OM research space as captured by topics and methods combined.

In conclusion, we believe that our application study has demonstrated the value of LSA as a method for surfacing the intellectual structure of the OM discipline. The applicability of LSA is not limited to OM. Any other discipline that falls within the Decision Sciences umbrella can be treated following the steps outlined in this study.

DIRECTIONS FOR FUTURE APPLICATIONS OF LSA

As the present article has demonstrated, LSA is a powerful text analysis method for distilling latent concepts. It may well lend itself to better understanding of a field of intellectual inquiry and, thereby, more informed decision making. Because LSA essentially relies on statistical analysis of words used in a set of documents, the method can be used in a number of settings that allow access to a reasonably large number of textual documents. For example, LSA can be used as part of a case study where the researcher can collect and analyze textual data for a number of observations to better understand patterns and connections between operational actions and performance outcomes that could be hard to recognize through manual analysis. We provide, below, some specific examples of potential LSA applications in OM research.

Transportation Safety

Safety at road construction sites has been a major concern for both transportation agencies and highway contractors. To coordinate efforts in addressing such concerns, the U.S. Department of Transportation's (DOT) National Highway Transportation Safety Administration (NHTSA) has established the Work Zone Mobility and Safety Program (http://www.ops.fhwa.dot.gov/wz/). Various regional initiatives exist around the country. For example, the New York State DOT Construction Accident Reporting Program requires contractors to provide accident data on worker injuries occurring during highway work activities and is considered one of the most comprehensive in the United States. Information on types of injuries and narratives as to the events leading up to the accidents have been useful in postaccident reviews of trends by the agency. Other agencies may eventually choose to adopt a similar data collection and assessment approach. LSA can be useful with the monitoring and systematic improvement of work zone safety within and across states by providing an efficient and standardized way to process incident narratives.

In another initiative related to transportation safety, NHTSA regularly receives vehicle product safety complaints through its Vehicle Owner Questionnaires (http://www.safercar.gov). The complaint information, which is voluntarily provided by consumers (who often lack expertise in automotive technology), is entered into the NHTSA consumer complaint database and catalogued according to vehicle or equipment make, model, model year, and the affected part (component description), assembly, or system as identified by the consumer. Currently, NHTSA technical staff read each complaint as it is received as part of a continuous review to identify potential trends that may indicate the presence of an emerging safety defect. The same data are also used to support existing safety defect investigations. LSA can be used to automate the processing of defect descriptions in order to provide for more efficient identification of trends in safety issues.

Acquisition of Products and Services

Acquisition of products and services in the U.S. Department of Defense (DoD) has almost doubled during the past decade, with about $387 billion spent in 2011 (Government Accounting Office, 2012). Given this significant expenditure of tax-payer dollars, it is important that the DoD selects the right contractor in its purchasing and contracting actions (Rendon, Apte, & Apte, 2012). One critical tool used by the DoD for this purpose is the Contractor Performance Assessment Reporting System (CPARS), a centralized data repository of contractor performance information. The CPARS requires that in addition to providing adjectival ratings, from exceptional to unsatisfactory, on performance dimensions of cost, quality, and schedule, the assessing official also provides a brief, factual narrative about the contractor's performance. In many cases, this narrative provides a rich source of information that proves to be very valuable in awarding best value contracts and orders to contractors that consistently provide quality, on-time products and services that conform to contractual requirements. As a tool for text analysis, the LSA methodology can prove to be very useful in this regard.

Healthcare Operations

Clinical decision making typically involves the use of both structured and unstructured (free-text) data. These data could include patient records (Acharya et al., 2013), unstructured interviews (Vosbergen et al., 2013), and transcripts of hospital communications (Pelayo, Anceaux, Rogalski, Elkin, & Beuscart-Zephir, 2013). While the structured data such as the specific diagnosis are obligatory, the free text is also an important part of patient records reflecting the thought process of the physician and other care givers. Despite the inherent value of the clinical information contained within a free-text record, its manual review can be a time-consuming activity. Hence, there exists great interest in developing approaches for extracting and analyzing such information from patient records (Behara, Fatteh, Rajadesingh, Jain, & Agarwal, 2013). In addition to being time consuming, this is also a complex task due to the ambiguity and variety of language used in the description and evaluation of patient conditions as well as the context and the use of terminology and acronyms. It is in this context that LSA can be valuable for analyzing patients' clinical records and initiating suitable interventions for providing long-term, high-quality healthcare to the patients. However, we caution the reader that healthcare-related data may be harder to obtain and should be handled in accordance with applicable law and with the approval of the appropriate institutional review board (IRB).

In summary, there are several avenues for using LSA in future research, besides uncovering the intellectual core of a field. What we have presented above is but a small sample of potential applications, and it is clear that scholars can independently and creatively address any area where the analysis of unstructured text data can provide vital insights. To that end, we hope that our work in this article generates interest in LSA and spurs research activity using this powerful methodological tool.

REFERENCES

Acharya, A., Hernandez, P., Thyvalikakath, T., Ye, H., Song, M., & Schleyerb, T. (2013). Development and initial validation of a content taxonomy for patient records in general dentistry. *International Journal of Medical Informatics*, *82*, 1171–1182.

Amoako-Gyampah, K., & Meredith, J. R. (1989). The operations management research agenda: An update. *Journal of Operations Management*, *8*(3), 250–262.

Applebaum, W. (1947). The Journal of Marketing: The first ten years. *Journal of Marketing*, *11*(4), 355–363.

Banker, R. D., & Kauffman, R. J. (2004). The evolution of research on information systems: A fiftieth-year survey of the literature in Management Science. *Management Science*, *50*(3), 281–298.

Behara, R., Fatteh, F., Rajadesingh, P., Jain, P., & Agarwal, A. (2013). A prognostic COPD clinical support system. *2013 International IEEE Systems Conference*, Orlando, FL.

Berry, M. W., Dumais, S., & O'Brien, G. (1995). Using linear algebra for intelligent information retrieval. *SIAM Review*, *37*(4), 573–595.

Boulding, K. E. (1958). Evidences for an administrative science: A review of the Administrative Science Quarterly, volumes 1 and 2. *Administrative Science Quarterly*, *3*(1), 1–22.

Bradford, R. B. (2008). An empirical study of required dimensionality for large-scale latent semantic indexing applications. *CIKM '08: Proceeding of the 17th ACM Conference on Information and Knowledge Management*. New York, NY: ACM, 153–162.

Brutus, S., Aguinis H., & Wassmer U. (2013). Self-reported limitations and future directions in scholarly reports: Analysis and recommendations. *Journal of Management*, *39*(1), 48–75.

Chopra, S., Lovejoy, W., & Yano, C. (2004). Five decades of operations management and the prospects ahead. *Management Science*, *50*(1), 8–14.

Cios, K. J., Pedrycz, W., Swiniarski, R. W., & Kurgan, L. A. (2007). *Data mining: A knowledge discovery approach*. New York, NY: Springer.

Craighead, C. W., & Meredith, J. R. (2008). Operations management research: Evolution and alternate future paths. *International Journal of Operations & Production Management*, *28*(8), 710–726.

Culnan, M. J. (1986). The intellectual development of Management Information Systems, 1972–1982: A co-citation analysis. *Management Science*, *32*(2), 156–172.

Deerwester, S., Dumais, S., Furnas, G., Landauer, T., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.

Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology*, *38*, 189–230.

Dumais, S. T. (2007). LSA and information retrieval: Getting back to basics. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, 293–322.

Efron, M. (2005). Eigenvalue-based model selection during latent semantic indexing. *Journal of the American Society for Information Science and Technology*, *56*(9), 969–988.

Evangelopoulos, N., Zhang, X., & Prybutok, V. (2012). Latent semantic analysis: Five methodological recommendations. *European Journal of Information Systems*, *21*(1), 70–86.

Filippini, R. (1997). Operations management research: Some reflections on evolution, models and empirical studies in OM. *International Journal of Operations & Production Management*, *17*(7), 655–670.

Government Accountability Office. (2012, March). *Defense contracting: Competition for services and recent initiatives to increase competitive procurements* (GAO-12–384). Washington, DC: GAO. [Accessed July 5, 2013, available at http://www.dtic.mil/cgibin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA558131.]

Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques (2nd ed.)*. San Francisco, CA: Morgan Kaufmann (Elsevier).

Hoffman, D. L., & Holbrook, M. B. (1993). The intellectual structure of consumer research: A bibliometric study of author cocitations in the first 15 years of the Journal of Consumer Research. *Journal of Consumer Research*, *19*, 505–517.

Kassarjian, H. H. (1977). Content analysis in consumer research. *Journal of Consumer Research*, *4*(1), 8–18.

Landauer, T. K. (2007). LSA as a theory of meaning. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, 3–32.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Manning C., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. New York, NY: Cambridge University Press.

Martin, D. I., & Berry, M. W. (2007). Mathematical foundations behind latent semantic analysis. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, 33–57.

Meredith, J. R., Raturi, A., Amoako-Gyanpah, K., & Kaplan, B. (1989). Alternative research paradigms in operations. *Journal of Operations Management*, *8*(4), 297–326.

Nerur, S. P., Rasheed, A. A., & Natarajan, V. (2008). The intellectual structure of the strategic management field: An author co-citation analysis. *Strategic Management Journal*, *29*, 319–336.

Pannirselvam, G. P., Ferguson, L. A., Ash, R. C., & Siferd, S. P. (1999). Operations management research: An update for the 1990s. *Journal of Operations Management*, *18*(1), 95–112.

Pelayo, S., Anceaux, F., Rogalski, J., Elkin, P., & Beuscart-Zephir, M. C. (2013). A comparison of the impact of CPOE implementation and organizational determinants on doctor–nurse communications and cooperation. *International Journal of Medical Informatics*, *82*, e321–e330.

Pilkington, A., & Fitzgerald, R. (2006). Operations management themes, concepts and relationships: A forward retrospective of IJOPM. *International Journal of Operations & Production Management*, *26*(11), 1255–1275.

Pilkington, A., & Liston-Heyes, C. (1999). Is production and operations management a discipline? A citation/co-citation study. *International Journal of Operations & Production Management*, *19*(1), 7–20.

Pilkington, A., & Meredith, J. (2009). The evolution of the intellectual structure of operations management—1980–2006: A citation/co-citation analysis. *Journal of Operations Management*, *27*, 185–202.

Porter, M. (1980). An algorithm for suffix stripping. *Program*, *14*(3), 130–37. Republished as: Porter, M. (2006). An algorithm for suffix stripping. *Program: Electronic Library and Information Systems*, *40*(3), 211–218.

Rendon, R., Apte, A., & Apte, U. (2012). Services contracting in the DoD: A comparison of management practices in the Army, Navy, and Air Force. *Defense Acquisition Research Journal*, *19*(1), 3–32.

Robertson, S. (2004). Understanding inverse document frequency: On theoretical arguments for IDF. *Journal of Documentation*, *60*(5), 503–520.

Rungtusanthanam, M. J., Choi, T. Y., Hollingworth, D. G., Wu, A., & Forza, C. (2003). Survey research in operations management: Historical analyses. *Journal of Operations Management*, *21*(4), 475–488.

Salton, G. (1975). A vector space model for automatic indexing. *Communications of the ACM*, *18*(11), 613–620.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, *24*, 513–523.

Scudder, G. D., & Hill, C. A. (1998). A review and classification of empirical research in operations management. *Journal of Operations Management*, *16*(1), 91–101.

Sidorova, A., Evangelopoulos, N., Valacich, J. S., & Ramakrishnan, T. (2008). Uncovering the intellectual core of the information systems discipline. *MIS Quarterly*, *32*(3), 467–482 & A1–A20.

Tsai, W., & Wu, C. (2010). Knowledge combination: A cocitation analysis. *Academy of Management Journal*, *53*(3), 441–450.

Valle-Lisboa, J. C., & Mizraji, E. (2007). The uncovering of hidden structures by latent semantic analysis. *Information Sciences*, *177*(19), 4122–4147.

Vessey, I., Ramesh, V., & Glass, R. (2002). Research in information systems: An empirical study of diversity in the discipline and its journals. *Journal of Management Information Systems*, *19*(2), 129–174.

Vosbergen, S., Janzen, J., Stappers, P. J., van Zwieten, M., Lacroix, J., Idema, K. et al. (2013). A qualitative participatory study to identify experiences of coronary heart disease patients to support the development of online self-management services. *International Journal of Medical Informatics*, *82*, 1183–1194.

Wei C-P, Hu P. J-H, Tai C-H, Huang C-N, & Yang C-S. (2008). Managing word mismatch problems in information retrieval: A topic-based query expansion approach. *Journal of Management Information Systems*, *24*(3), 269–295.

White, H., & McCain, K. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science*, *49*(4), 327–355.

Zhu, M., & Ghodsi, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics & Data Analysis*, *51*(2), 918–930.

Shailesh Kulkarni is an associate professor of decision sciences in the Information Technology and Decision Sciences department, at the University of North Texas (UNT). He is also a past *UNT Leadership Fellow*. He holds a BEng degree in civil engineering and a Master of Mgmt. Studies from the University of Mumbai, India, and an MS in quantitative analysis and PhD in operations management from the University of Cincinnati. His research interests include supply chain risk, manufacturing/service operations, and data mining. His publications include research articles appearing in *Production and Operations Management*, *IIE Transactions*, *European Journal of Operational Research*, and *International*

Journal of Production Research among others. His consulting experience includes corporate projects in supply chain modeling, call center operations, and complex logistics systems. In 2009 he received the inaugural UNT College of Business Outstanding Teaching Innovation Award. He currently serves on the board of the *Production and Operations Management Society* (POMS) and as their *Associate V.P. for Sponsorships*.

Uday Apte is a professor of operations management at the Graduate School of Business and Public Policy (GSBPP), Naval Postgraduate School, Monterey, CA. He also serves as the Associate Dean of Research and Development at GSBPP. Before joining NPS, he taught at the Wharton School, University of Pennsylvania, and at the Cox School of Business, Southern Methodist University. He holds a PhD in Decision Sciences from the Wharton School, University of Pennsylvania, MBA from the Asian Institute of Management, Manila, Philippines, and Bachelor of Technology (Chemical Engineering) from the Indian Institute of Technology, Bombay, India. His current research interests include managing operations in information economy with specific emphasis on information-intensive services, service operations, and supply chain management. He has published over 60 articles, five of which have won awards from professional societies. His research articles have been published in prestigious journals including *Management Science, Production and Operations Management, Journal of Operations Management, Interfaces, IIE Transactions,* and *MIS Quarterly*. He has coauthored two books, *Manufacturing Automation* and *Managing in the Information Economy*. He has served as a vice president of colleges at Production and Operations Management Society (POMS), and as a founder and president of the POMS College of Service Operations. Prior to his career in academia, he worked for over 10 years in managing operations and information systems in the financial services and utility industries. Since then he has consulted with several major U.S. corporations and international organizations.

Nicholas Evangelopoulos is an associate professor of decision sciences in the Information Technology and Decision Sciences department, at the University of North Texas (UNT), and a Fellow of the Texas Center for Digital Knowledge. He holds a BEng degree in electrical engineering from Aristotle University of Thessaloniki, Greece, an MS in computer science from the University of Kansas, and a PhD in decision sciences from Washington State University. His research interests include applied statistics and text analytics. His publications include research articles appearing in *MIS Quarterly*, *Decision Support Systems*, *Communications of the ACM*, *Communications in Statistics*, *Computational Statistics & Data Analysis*, and many others. His consulting experience includes corporate projects in text analytics, predictive modeling, customer attrition modeling, and policy evaluation. In 2010 he received the UNT College of Business Outstanding Teaching Innovation Award and in 2013 the UNT College of Business Outstanding Senior Faculty Research Award.